Graphical Report

IM921: Visualisation

Estefanía Magaña
U1856249

Fig. 1

I. Introduction

We currently live in a world that drowns us with information. However, being able to extract knowledge from those data resources is not always an easy task. Data visualisation plays a vital role in statistical data analysis, and represents an essential part in closing the gap between having thousands of figures and provide answers to relevant questions [1].
Data visualisation represent a non-verbal mean to express information even on its higher complexity level (or not) into a simpler and more straightforward way. Therefore, visualisation is a central piece in the discovery, understanding, and communication of information in many fields [4] [6].
In this work we focus in the storytelling ability of data visualisation. A narrative visualisation with data, contrary with the ways stories are presented in text, can be organized in linear and non-linear sequence, provide alternative explanations, and even generate new questions [3]. In the attempt to produce a story telling graph, we kept in mind the following primary goals of visualisation: accuracy, accessibility an aesthetical appeal [8]. Yet, in order to gain aesthetical appeal, we allowed ourselves to relax the contextualization of the data and we lost some comparison enforcement. However, these factors are not exclusive. In fact, the plurality of options on the design spectrum is able to improve the effectiveness of the knowledge extraction while visualising data [6].
In the visualisation displayed in Figure 1 we looked to generate an attractive to the eye data-driven design. As inspiration, we used the Dominoes piece of Gregory Matthews [7] displayed in Figure 2. While taking this image as our base, we decided to bring in small multiples power to present multivariate information as our set of tools to present the data.

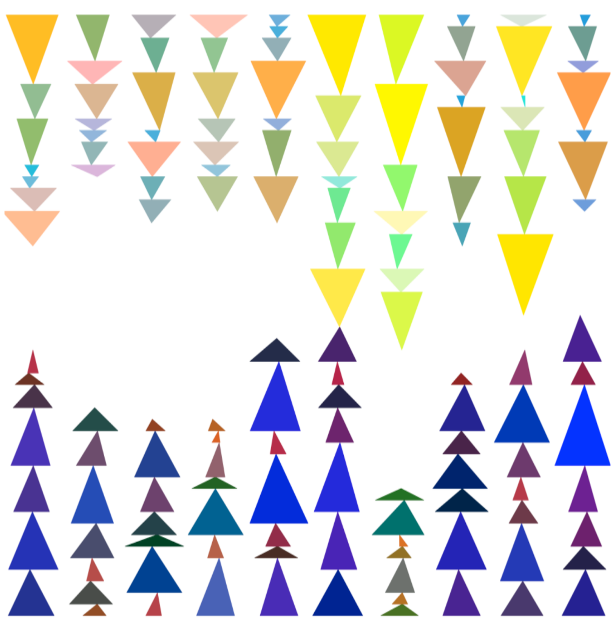

Fig. 2: Dominoes. Source: Matthews (2018) [7]

The rest of this paper begins with a brief explanation of the dataset used in this project, as well as the software tools employed in order to create the graph displayed in Figure 1. We continue by explaining the ideas behind this visualisation and end up with our main findings during this journey. A detailed bibliography is presented at the end of the paper.

II. Data & Software

The dataset underlying Figure 1 is about The Spanish Armada, composed 130 ships distributed in 10 fleets. Besides the name of the fleet and the number of ships on each one of them, the dataset presents additional numeric information of each of these fleets making a data frame with ten observations and 11 variables in total (one categoric and ten numeric). An extract of the dataset is provided in Table I.

TABLE I: User-Item Rating Matrix for a Movie Recommendations Application

Armada	ships	tons	soldiers	sailors	···	rope
Portugal	12	7737	3330	1293		150
Vizca	14	6567	1937	863		87
Castilla	16	8714	2458	1719		309
⋮	⋮	⋮	⋮	⋮	⋱	⋮
Galeras	4	0	0	362	···	20

The statistical computing software R [9] is the main tool used in this work to handle the data and to generate the visual design showed in Fig 1. The two packages we employed were *HistData* [2] and *showtext* [5]. The full R code to recreate the visualization displayed in Figure 1 is detailed in Appendix A. Additionally, even though the *wesanderson* [10] package is not explicitly used, we implemented the colour palette in shown in Figure 3.

Fig. 3: "Isle of Dogs 2" Colour Palette. Source: Ram and Wickham (2015) [10]

III. The Graph

The starting point of sketching our dataset was to find out how to arrange the information. First, we had to be able to compare all the different variables. As previously explained, the dataset has ten numeric variables, all of which have a significant variation of scales. To avoid giving more emphasis to those variables with the grater numeric, we decided to rescale the normalise all variables within the range 1 to 2.
We continued by playing with different aspects of the scene, such as the distribution of the fleets and the arrangement of the features. We used color and size to show the ranking of

the elements. As the data was normalised into a low difference range, using these two features will allow easier comparison. Together, they form an imperfect visual description.

Figure 1 presents a particular way to play with shapes and spaces. A random element was added to the shapes of the triangles to give the design a "sketchier" look; this avoids perfect and equal forms. The way the paths were arranged has an inconstant factor. While linear ordering was used for the different fleets, there are elements stacked from bottom to top and other from top to bottom. On this light, we are aware of the limitation of a direct comparison, as we are biasing the scope of the eye perception.

The usage of all the displayed space gives its part to the graphical design, which only includes annotations on the horizontal coordinate. The primary objective was to let the image gain the main attention, then draw the eyes of the viewer into the title. This title was selected to present the overall intention of the design, which was a sketch with a data set. Of course, a significant drawback is to lose context and therefore the dataset is no longer the main character. However, our narrative is not the Armada fleets, but a way to show a sketch storytelling visualisation in a personal style.

Alternatively, we could have chosen to represent every fleet feature on a bar plot arranged in different facets, if the objective was to compare one against each other. Another option would be to present a bar plot per fleet (all the features in the x-axis) with the mean, or median level displayed as a line. These data presentation would be an informative way to visualise the composition of each fleet compared to the overall structure of the Spanish Armada. Then again, the data set was not the central pease of this work's visual representation.

IV. Conclusion

Data visualisation is a central feature in many fields that leads to discovery, understanding and communication. In our graphical journey we were even more aware of all the decision that play a part into visually communicate information. However, there are multiple mistakes that can lead to poor clarity and even sending the wrong message.

By bringing together a narrative visualisation with small multiples tools, we were able to create an appealing data-driven design. In this design, the viewer can observe the illustration on its own. Then, with the additional information provided by the labels and annotation, the viewer can explore the design in a more informative way.

By focusing on the appealing side of the graphic, we gave less attention to the contextualization and clear comparison of the dataset. Given the importance of these elements, we recognize the limitations of the visualisation presented in this work, as it does not clearly describe the dataset behind it.

References

[1] Cleveland, W.S. and McGill, R., 1984. "Graphical perception: Theory, experimentation, and application to the development of graphical methods". *Journal of the American statistical association*, 79(387), pp.531-554.

[2] Friendly, M., Dray, S., Wickham, H., Hanley, J., Murphy, D. and Li, P., 2018. Package 'HistData'.

[3] Segel, E. and Heer, J., 2010. Narrative visualization: Telling stories with data. IEEE transactions on visualization and computer graphics, 16(6), pp.1139-1148.

[4] Healy, K. and Moody, J., 2014. "Data visualization in sociology". *Annual review of sociology*, 40, pp.105-128.

[5] Qiu, Y., 2019. Package 'showtext'.

[6] Kurlansky, M., Marcus, S., Reineck, J., Reineck, G. and Chairman-Marcus, A., 1980, July. "Special session on effective information presentation techniques". In *ACM SIGGRAPH Computer Graphics* (Vol. 14, No. 3, p. 1). ACM.

[7] Matthews, G., 2015. *Stats in the wild*, [Online] accessed 4 March 2019, https://statsinthewild.com/art/

[8] Murray, S., 2015. "Changing Minds to Changing the World". In *New Challenges for Data Design* (pp. 293-312). Springer, London.

[9] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[10] Ram, K. and Wickham, H., 2015. "wesanderson: A Wes Anderson Palette Generator. R package version 0.3. 2."

APPENDIX

A. R code

```
# ================================
# SKETCHING WITH SMALL MULTIPLES
# ================================



library(HistData)
library(showtext)

# Data Processing

data(Armada)
Armada <- Armada[order(Armada$artillery),]

vars.int <- setdiff(names(Armada), c('Armada'))


# Normalisation [1,2]
for(i in vars.int){
  Armada[,paste(i,'_n', sep='')] <- (Armada[,i] - min(Armada[,i]))/
(max(Armada[,i]) - min(Armada[,i])) + 1

}

# Ranking each variable
for(i in vars.int){
  Armada[,paste(i,'_r', sep='')] <- round(rank(Armada[,paste(i,'_n', sep='')]))
}


set.seed(1234) # to make it replicable

# Colour Selection

FTbg <- rgb( 254, 241, 224, maxColorValue =255 ) # background colour
FTtx <- rgb( 114, 108, 97, maxColorValue =255 ) # text colour

pal <- c("#EAD3BF", "#CDB7A5", "#B19B8C", "#AE8F73", "#B38859", "#9A7246",
"#624D39", "#352E2C", "#282222", "#1C1718")

showtext_auto()

font_add_google('Indie Flower', 'indie')
font_add_google('Gochi Hand', 'gochi')

x11()

back <- '#FAEBD7'

par( bg = back,
     mar=c(0.1,0.1,0.1,0.1),  # the plot margins
     oma=c( 15,0,0,0 ) )      # outer marginsbottom


# Colour Palette

top <-14




plot(-99,-99, xlim=c(0,29), ylim=c(0,top), ann=F, axes=F, type='n', xaxs='i', yaxs='i')


for(i in seq(0,nrow(Armada)-1,1)){

  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(0,Armada$ships_n[i+1],0),
          col = pal[Armada$ships_r[i+1]], border = NA)
  extra <- Armada$ships_n[i+1]
```

```
  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(extra,Armada$soldiers_n[i+1]+extra,extra),
          col = pal[Armada$soldiers_r[i+1]], border = NA)
  extra<- extra+Armada$soldiers_n[i+1]
  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(extra,Armada$sailors_n[i+1]+extra,extra),
          col = pal[Armada$sailors_r[i+1]], border = NA)

}


war <- c('artillery_n','balls_n','gunpowder_n','lead_n','rope_n')
war.r <- c('artillery_r','balls_r','gunpowder_r','lead_r','rope_r')

for(i in seq(0,nrow(Armada)-1,1)){

  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(top,top-Armada[i+1,war[1]],top),
          col = pal[Armada[i+1,war.r[1]]], border = NA)
  extra <- Armada[i+1,war[1]]
  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(top-extra,top-extra-Armada[i+1,war[2]],top-extra),
          col = pal[Armada[i+1,war.r[2]]], border = NA)
  extra <- extra+Armada[i+1,war[2]]
  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(top-extra,top-extra-Armada[i+1,war[3]],top-extra),
          col = pal[Armada[i+1,war.r[3]]], border = NA)
  extra <- extra+Armada[i+1,war[3]]
  polygon(c(i*3,i*3+runif(1,0.5,1.5),2+i*3),
          c(top-extra,top-extra-Armada[i+1,war[4]],top-extra),
          col = pal[Armada[i+1,war.r[4]]], border = NA)

}
op <- par(family = "indie")
axis(side=1, seq(1,28,3),labels=c('N','Ga','U','A','P','Vi','Gu','C','P','Vr'), col=NA)



# =============
#   SUBTITLES
# =============

op <- par(family = "gochi")
title( sub = "Sketching with Small Multiples" ,
       cex.sub =2,
       outer=T, adj=0, line=3)

title( sub = "La Gran y Felicisima Armada" ,
       cex.sub =1,
       outer=T, adj=0, line=4.2)



# ===========
#   LEGENDS
# ===========


op <- par(family = "indie")


legend(-1,-3.7, bty='n', xpd=NA, title='Fleet:',
       legend=c('N:  Napoles','Ga: Galeras','U:  Uantiscas','A:  Andalucia','P:  Portugal'),
       cex=1.3)
legend(5,-4.6, bty='n', xpd=NA,
       legend=c('Vi: Vizca','Gu: Guipuscua','C:  Castilla','P:  Pataches','Vr: Vrcas'),
       cex=1.3)


# All of these could be inserted in a loop... sorry to leave it explicit

# COLOR AND SIZE LEGEND

x <- 13.8
```

```
y <- -3.7

legend(x,y, bty='n', xpd=NA, title='Ranking*:',
       legend=rev(c('','','','','')),
       col=rev(pal[6:10]),
       pch=c(NA,NA,NA,NA,NA), cex=1.3)
legend(x,y, bty='n', xpd=NA, title='',
       legend=rev(c('6th','','','','')),
       col=rev(pal[6:10]),
       pch=c(NA,NA,NA,NA,17), cex=1.3, pt.cex=1.99)
legend(x,y, bty='n', xpd=NA, title='',
       legend=rev(c('','7th','','','')),
       col=rev(pal[6:10]),
       pch=c(NA,NA,NA,17,NA), cex=1.3, pt.cex=2.09)
legend(x,y, bty='n', xpd=NA, title='',
       legend=rev(c('','','8th','','')),
       col=rev(pal[6:10]),
       pch=c(NA,NA,17,NA,NA), cex=1.3, pt.cex=2.19)
legend(x,y, bty='n', xpd=NA, title='',
       legend=rev(c('','','','9th','')),
       col=rev(pal[6:10]),
       pch=c(NA,17,NA,NA,NA), cex=1.3, pt.cex=2.29)
legend(x,y, bty='n', xpd=NA, title='',
       legend=rev(c('','','','','Max')),
       col=rev(pal[6:10]),
       pch=c(17,NA,NA,NA,NA), cex=1.3, pt.cex=2.39)

# min - 5

x <- 17.9
y <- -4.6

legend(x,y, bty='n', xpd=NA,
       legend=rev(c('Min','','','','')),
       col=rev(pal[1:5]),
       pch=c(NA,NA,NA,NA,17), cex=1.3, pt.cex=1.49)
legend(x,y, bty='n', xpd=NA,
       legend=rev(c('','2nd','','','')),
       col=rev(pal[1:5]),
       pch=c(NA,NA,NA,17,NA), cex=1.3, pt.cex=1.59)
legend(x,y, bty='n', xpd=NA,
       legend=rev(c('','','3th','','')),
       col=rev(pal[1:5]),
       pch=c(NA,NA,17,NA,NA), cex=1.3, pt.cex=1.69)
legend(x,y, bty='n', xpd=NA,
       legend=rev(c('','','','4th','')),
       col=rev(pal[1:5]),
       pch=c(NA,17,NA,NA,NA), cex=1.3, pt.cex=1.79)
legend(x,y, bty='n', xpd=NA,
       legend=rev(c('','','','','5th')),
       col=rev(pal[1:5]),
       pch=c(17,NA,NA,NA,NA), cex=1.3, pt.cex=1.89)


# VERTICAL DISTRIBUTION EXPLANATION

legend(22,-2.2, bty='n', xpd=NA, title='Feature Distribution:',
       legend=c("Artillery","Balls","Gunpowder","Lead","Rope"," ", "Sailor","Soldier","Ships"),
       pch=c(6,6,6,6,6,NA,2,2,2),
       cex=1)

# Some further notes

title( sub = "*Rankings were obtained by normalising each feature to values within [1,2]",
       col = FTbg ,
       cex.sub=0.8,
       outer=T, adj=0, line=13.8)

title( sub = "Source: HistData",
       col = FTbg ,
       cex.sub=0.8,
       outer=T, adj=1,line=14)

dev.off()
```